

# UC Berkeley

## UC Berkeley Previously Published Works

**Title**

Suboptimal eye movements for seeing fine details.

**Permalink**

<https://escholarship.org/uc/item/8zv8c078>

**Journal**

Journal of vision, 18(5)

**ISSN**

1534-7362

**Authors**

Agaoglu, Mehmet N  
Sheehy, Christy K  
Tiruveedhula, Pavan  
et al.

**Publication Date**

2018-05-01

**DOI**

10.1167/18.5.8

Peer reviewed

# Suboptimal eye movements for seeing fine details

**Mehmet N. Ağaoğlu**

School of Optometry, University of California, Berkeley,  
Berkeley, CA, USA  
Vision Science Graduate Group, University of California,  
Berkeley, Berkeley, CA, USA



**Christy K. Sheehy**

Department of Neurology, University of California,  
San Francisco, San Francisco, CA, USA



**Pavan Tiruveedhula**

School of Optometry, University of California, Berkeley,  
Berkeley, CA, USA



**Austin Roorda**

School of Optometry, University of California, Berkeley,  
Berkeley, CA, USA  
Vision Science Graduate Group, University of California,  
Berkeley, Berkeley, CA, USA



**Susana T. L. Chung**

School of Optometry, University of California, Berkeley,  
Berkeley, CA, USA  
Vision Science Graduate Group, University of California,  
Berkeley, Berkeley, CA, USA



**Human eyes are never stable, even during attempts of maintaining gaze on a visual target. Considering transient response characteristics of retinal ganglion cells, a certain amount of motion of the eyes is required to efficiently encode information and to prevent neural adaptation. However, excessive motion of the eyes leads to insufficient exposure to the stimuli, which creates blur and reduces visual acuity. Normal miniature eye movements fall in between these extremes, but it is unclear if they are optimally tuned for seeing fine spatial details. We used a state-of-the-art retinal imaging technique with eye tracking to address this question. We sought to determine the optimal gain (stimulus/eye motion ratio) that corresponds to maximum performance in an orientation-discrimination task performed at the fovea. We found that miniature eye movements are tuned but may not be optimal for seeing fine spatial details.**

## Introduction

We make large, rapid, voluntary eye movements—saccades—to redirect our gaze to accomplish numerous visual tasks (e.g., searching for an object, reading a

book, etc.; Kowler, 2011). This is to form a fine-grained representation of the external world by taking advantage of a part of the retina—the fovea—which has the highest spatial resolution. However, our eyes are always in motion between epochs of saccades, even when we try to maintain our gaze on an object. Miniature eye movements that we make during fixation are often referred to as fixational eye movements (FEM). Different types of FEM have been identified depending on their spatiotemporal characteristics (Martinez-Conde, Macknik, & Hubel, 2004; Rucci & Poletti, 2015). Microsaccades are small jerky eye movements and share similar peak velocity–amplitude dynamics as larger saccades (Zuber, Stark, & Cook, 1965). Drifts are relatively slower and smoother but rather erratic eye movements that occur between (micro)saccades and have been usually modeled as various types of random walk or Brownian motion (Burak, Rokni, Meister, & Sompolinsky, 2010; Engbert, Mergenthaler, Sinn, & Pikovsky, 2011; Rucci, Iovin, Poletti, & Santini, 2007). Lastly, tremors are usually defined as very low-amplitude and high-frequency oscillatory movements that are superimposed on drifts (Ditchburn & Ginsborg, 1953; H. K. Ko, Snodderly, & Poletti, 2016; Ratliff & Riggs, 1950).

Citation: Ağaoğlu, M. N., Sheehy, C. K., Tiruveedhula, P., Roorda, A., & Chung, S. T. L. (2018). Suboptimal eye movements for seeing fine details. *Journal of Vision*, 18(5):8, 1–20, <https://doi.org/10.1167/18.5.8>.



In addition to the nonuniform distribution of density and size of receptive fields of ganglion cells across the retina (Curcio, Sloan, Kalina, & Hendrickson, 1990; Dacey & Petersen, 1992; A. B. Watson, 2014), there is another unique property that differentiates our visual system from a computer vision system—neural adaptation. For instance, retinal ganglion cells (RGC) are most responsive to light transients, and their responses decay with prolonged exposure (Benardete & Kaplan, 1997; Kaplan & Benardete, 1999). Although such a system is ideal for the detection of changes or movements that are crucial for survival in a natural setting, it comes with a consequence. It has been known that, in the absence of FEM, visual perception fades away and more so for small visual stimuli (Ditchburn & Ginsborg, 1952; Riggs, Ratliff, Cornsweet, & Cornsweet, 1953; Yarbus, 1967). This suggests that retinal image motion is essential for continuous and high-acuity vision. However, if FEM are too large or too fast, the light intensity defining a visual stimulus will spread over a large population of cells, each with an insufficient exposure to the stimulus, resulting in nothing but smeared, ghost-like impressions. Therefore, there must be an optimum movement between these two extremes whereby visual perception is maximized such that the ability to perform a visual task is highest at that particular movement.

Previous research showed that naturally occurring or artificially induced irregular and continuous retinal image drifts help in seeing fine spatial details (Ratnam, Harmening, & Roorda, 2017; Rucci et al., 2007). Likewise, naturally occurring microsaccades or sudden jumps of stimuli are known to counteract visual fading (Costela, McCamy, Macknik, Otero-Millan, & Martinez-Conde, 2013; Martinez-Conde, Otero-Millan, & Macknik, 2013) and help redirect our gaze to compensate for the nonhomogeneous acuity within the fovea (Poletti, Listorti, & Rucci, 2013). If there is a causal relationship between FEM and visual perception, the latter should show a “tuning” function with different levels of the former, analogous to orientation tuning of cells in the early visual areas in which firing rate of a given cell is continuously modulated by how close the orientation of a stimulus in its receptive field is to its “preferred orientation.” In other words, the way a retinal image moves as a function of FEM, i.e., the ratio of the stimulus velocity and the actual eye velocity, should result in systematic changes in visual perception, and maximum performance in a visual task would be obtained at the preferred or optimal FEM. This velocity ratio was the independent variable in this study and is referred to as “gain” in the remainder of this article. Here, we asked observers to report the orientation of a grating presented at their fovea while undergoing several gain manipulations. Our results show that normal FEM are tuned but not quite optimal

for fine discrimination at the fovea: The best orientation discrimination performance occurred at a gain of  $\sim 0.4$  (rather than at a gain of zero, which represents retinal motion due to natural FEM). We also found that within the range of spatial frequencies in which the human visual system has highest contrast sensitivity, the tuning relationship between gain and performance disappears, suggesting a higher tolerance for retinal image motion for coarse visual structures.

## Methods

### Participants

Seven human subjects (including the first author, S1, ages 18–35) with normal or corrected-to-normal vision (20/20 or better in each eye) participated in the study. All seven subjects took part in Experiment 1. Three of the seven subjects participated in Experiment 2. All subjects except the first author were naïve as to the purpose and the details of the experiments. All subjects gave written informed consent prior to the experiments. All experimental procedures followed the principles put forth by the Declaration of Helsinki and were approved by the Institutional Review Board at the University of California, Berkeley.

### Apparatus

For stimulus delivery and eye tracking, we used a custom-built tracking scanning laser ophthalmoscope (TSLO; Sheehy et al., 2012). The TSLO has a diffraction-limited optical design; provides high-fidelity imaging of the retina; and more importantly, offers online tracking of eye movements. For the experiments presented here, we used a  $10 \times 10 \text{ deg}^2$  ( $512 \times 512$  pixels<sup>2</sup>) field of view (FOV), which yielded a pixel size of 1.17 arcmin. A large FOV enabled us to capture videos with rich retinal structure, which, in turn, allowed accurate image-based eye tracking and stimulus delivery. The horizontal scanner operates at 16 kHz whereas the vertical scanner operates at 1/512 of this rate to record full frames at 30 frames per second. An 840-nm super luminescent diode with a 50-nm bandwidth was used to scan the retina. Visual stimuli were delivered by manipulating the intensity of the laser beam via an acousto-optic modulator with the output controlled by a 14-bit digital-to-analog converter. Therefore, the stimuli had a negative contrast on the dim red raster created by the scanner (i.e., appeared black on a red background). Details of online eye-movement tracking have been reported elsewhere (Arathorn et al., 2007; Mulligan, 1998; Yang, Ara-

thorn, Tiruveedhula, Vogel, & Roorda, 2010). Briefly, each frame was broken into 32 horizontal strips ( $16 \times 512$  pixels<sup>2</sup>), and each strip was cross-correlated with a reference frame acquired earlier. The horizontal and vertical shifts required to match a strip to the reference frame represent a measure of the relative motion of the eye. This method results in an eye-movement sampling rate of 960 Hz. These computations occur in near real time (with  $2.5 \pm 0.5$  ms delay) and allows accurate stimulus delivery at specific retinal locations.

## Stimuli and procedures

The task was to report the orientation of a sinusoidal grating from vertical (two-alternative, forced choice; clockwise or counterclockwise). The amount of tilt was  $\pm 45^\circ$  from vertical, and the spatial frequency of the grating was 12 c/deg in Experiment 1 (Figure 1a) and 3 c/deg in Experiment 2. Prior to each experiment, the contrast of the grating was adjusted for each subject to yield  $\sim 80\%$  correct discrimination performance under the natural viewing condition (i.e., gain = 0). The size of the grating was  $3 \times 0.75$  deg<sup>2</sup>. Horizontal edges were smoothed using a cosine profile. Mean luminance (as measured indirectly from laser power) of the grating was kept at  $\sim 70\%$  of the luminance of the scanning raster for all subjects regardless of the contrast of the grating (Figure 1b). All experiments were performed in a dark room, and subjects were dark adapted for about half an hour prior to any data collection. For some subjects (three out of seven), the pupil of the imaged eye was dilated to maintain good retinal image quality throughout the session.

Each trial started with a fixation cross ( $0.2^\circ$ ) presented at the center of the raster. Subjects were asked to fixate at this cross before each trial. The experimenter manually acquired a reference frame for online tracking at the start of each trial. As soon as the experimenter initiated a trial, the fixation cross disappeared, and following a random delay (up to 1 s), the stimulus was presented at the center of the raster for 900 ms (flickered at 30 Hz). Subjects responded via a gamepad, had unlimited time to respond, and could have a break at any point during a block of trials. The independent variable was the ratio of the stimulus velocity on the raster and the actual eye velocity. Here, we refer to this as “gain.” A gain of zero means that stimulus stays stationary regardless of the eye movements. A gain of one represents perfect stabilization of the stimulus on the retina such that the ratio of stimulus motion and eye motion is one. The gain values between zero and one represent partial stabilization; values larger than one represent overstabilization (i.e., stimulus moves faster than the eye); and finally, values lower than zero amplify the retinal motion of the

stimulus by moving the stimulus in the opposite direction of the eye motion. The set of gains used were  $-1, -0.5, 0, 0.5, 1, 1.5$ , and 2. Different gains were interleaved within a block of trials. All observers completed at least 100 trials per gain. Right after the completion of the main experiment and before data analyses, subject S4 was tested for a second run in Experiment 1 because her retinal image quality and head stability during the experiment were poor, causing retinal stabilization to fail in many of her trials. In the second run, we used finer steps of gain (from  $-0.25$  to 1.25 in steps of 0.125), and subject S4 again ran at least 100 trials per gain. However, all results presented in the main text include only the data from the first run for S4. The results from both runs for S4 are shown in Figure 3.

## Retinal video analysis

Retinal videos were analyzed off-line for five main reasons. First, we sought to determine how well and where the stimulus was delivered on a trial-by-trial basis. Second, online eye tracking was performed by using raw retinal images, which were corrupted by high-frequency noise and low-frequency luminance gradients. In order to get more accurate eye motion estimates relatively less dependent on changes in overall brightness and uniformity of retinal images across trials, one needs to perform several preprocessing steps on retinal images. To this end, we performed the following image processing steps before computing eye motion: trimming, detection, and removal of frames during which subjects blinked, extracting stimulus position and removal of the stimulus (replaced by random noise patterns whose statistics—mean and standard deviation—matched to the rest of the frame), gamma correction, bandpass filtering (for removal of high-frequency noise and low-frequency brightness gradients), and making a reference frame. Third, during online eye tracking, if the peak of normalized cross-correlation between a strip and the reference frame was below 0.3, possibly due to (a) bad image quality, (b) excessive distortion of image features due to a rapid eye movement, (c) insufficient amount of overlap between the strip and the reference frame due to large eye motion, or (d) blinks, the stimulus was not delivered. By off-line processing of retinal videos, we also sought to inspect each and every frame of retinal videos and discard the trials if the stimuli was delivered inaccurately or was not delivered at all in more than two frames per trial (note that with this criterion, trials in which subjects blinked were also discarded). This procedure resulted in removal of 28.8% (1,305/4,525) and 8.7% (206/2,353) of all trials in Experiments 1 and 2, respectively. Fourth, because the reference frames

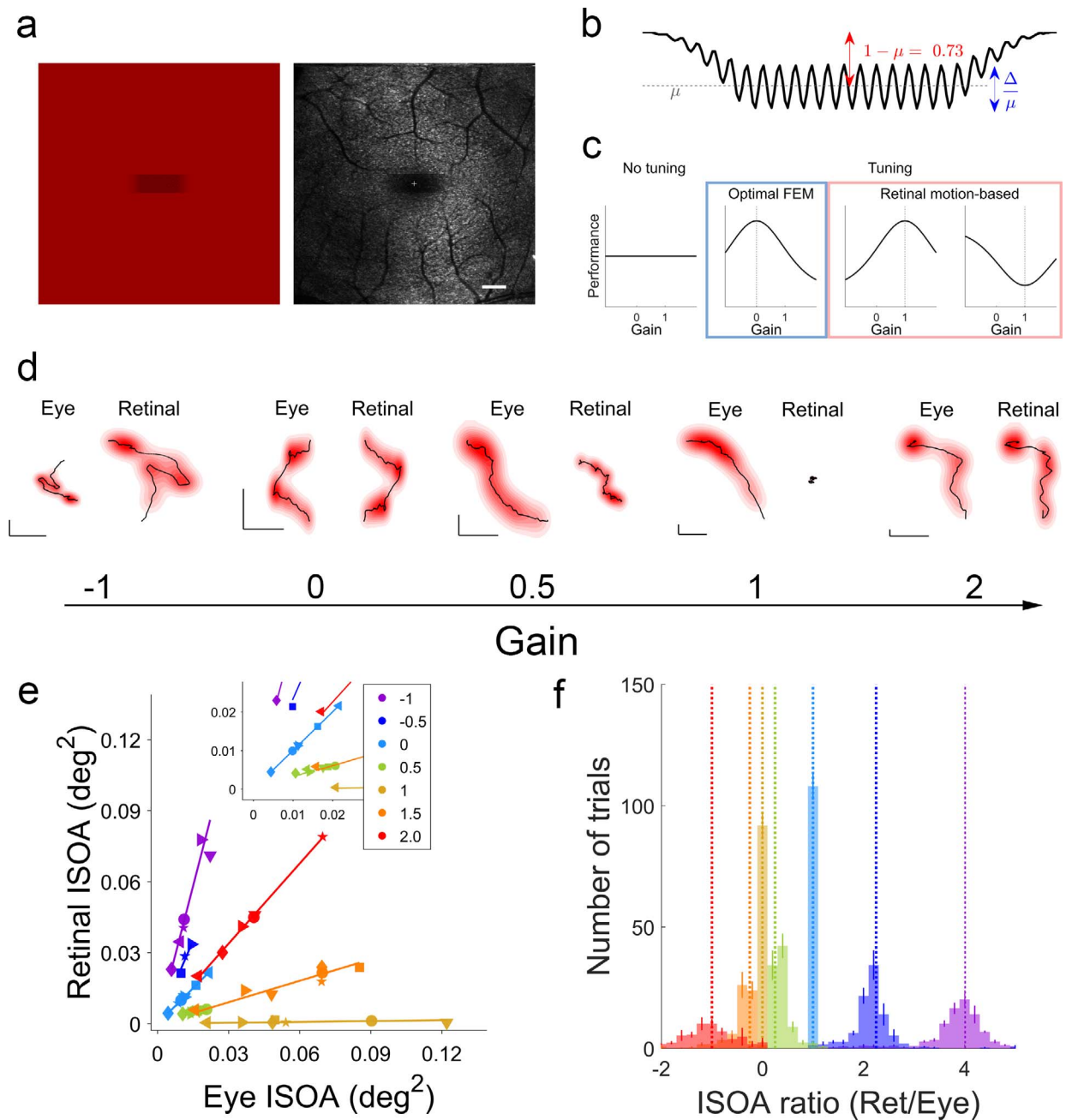


Figure 1. Manipulating the relationship between retinal image motion and eye motion with the TSLO. (a) An orientation discrimination task at the fovea. Subjects' view of a grating on the raster (left) and corresponding retinal image (right). Note that the stimulus is imprinted on the retinal image. (b) The luminance profile used to create grating patterns on the raster. The mean luminance of the grating was set to  $\sim 70\%$  of the background, and the contrast of grating was adjusted for each subject. (c) Predictions from the no-tuning (null) and tuning hypotheses. The panels with blue and red outlines show various ways tuning can occur. (d) Sample eye motion and retinal image motion traces (black lines) and corresponding probability densities (red clouds) for different gains. The horizontal and vertical lines in the lower left corner of each panel represent  $0.1^\circ$ . Dimensions were adjusted for clarity. (e) Retinal ISOA as a function of eye ISOA across gains in Experiment 1. Different colors represent different gains, and subjects are coded by different symbols. Inset shows a close-up view of data for smallest retinal/eye motion. (f) The distribution of retinal/eye ISOA ratios for different gains averaged across seven subjects. Vertical dotted lines show theoretical ISOA ratios, i.e., assuming that eye tracking, stimulus delivery, and offline eye movement extraction were perfect. Error bars represent  $\pm SEM$  ( $n = 7$ ). Color conventions for gains are identical across all figures.



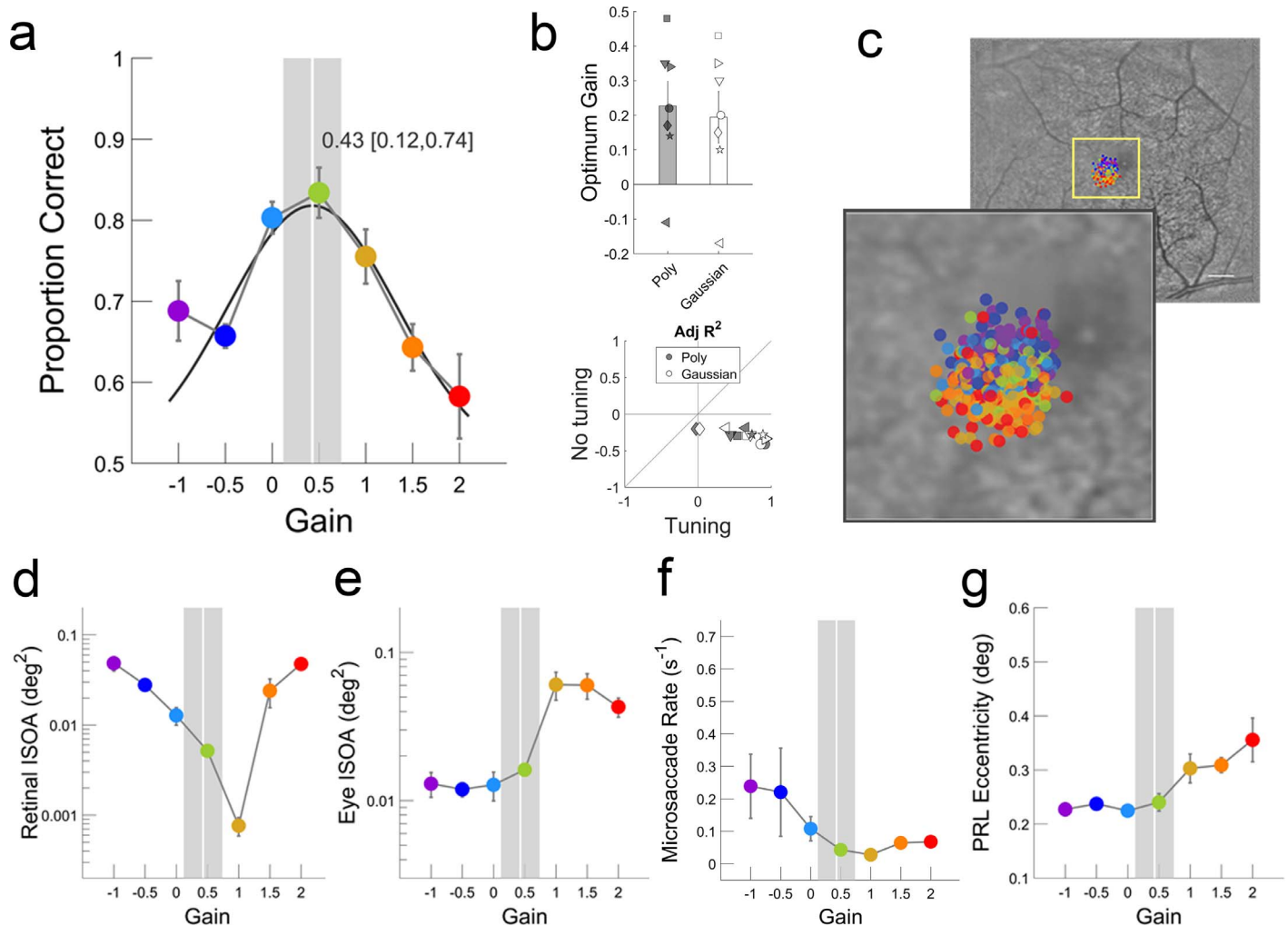


Figure 2. FEM are tuned but not optimal for fine discrimination at the fovea. (a) Proportion correct as a function of gain averaged across subjects in Experiment 1. A Gaussian tuning function was fit to all data (black curve) to estimate the optimum gain. Vertical white line represents optimal gain defined as the gain corresponding to the peak of the Gaussian. Shaded regions represent 95% confidence intervals of the optimum gain. The redundant color-coding here was necessary for Panel c. (b, top) Average optimal gains based on individual tuning function fits along with individual optimal gains. A quadratic polynomial and a Gaussian tuning function resulted in statistically indistinguishable optimal gains. (b, bottom) To compare the “no tuning” and “tuning” hypotheses in terms of how well they can explain our data, we computed Adjusted  $R^2$  metric for the constant model and tuning models (a quadratic polynomial or a Gaussian), respectively. For all subjects, tuning models performed better. (c, top right) The distribution of PRL across trials for one representative subject. Each symbol represents one trial. (c, bottom left) A close-up view of the central  $\sim 2.5^\circ$  part of the retina. Note the systematic change in PRLs across gains. (d, e) Retinal image motion and eye motion ISOA as a function of gain. (f, g) Microsaccade rates and PRL eccentricity across gains. Optimal gain and confidence intervals in panel a are replotted in Panels d through g. Error bars represent  $\pm SEM$  ( $n = 7$ ).

used for online tracking were basically snapshots of the retina taken manually by the experimenter and because the eyes are almost never stationary, the reference frames themselves might have some distortions due to these motions. By off-line processing, we created a relatively motion-free reference frame for each and every trial separately in an iterative process. This process started by selecting one of the frames in a retinal video as the reference frame, and computing eye motion. After each iteration, a new reference frame was

built by using computed eye motion and individual strips. We performed three iterations for each video, and the reference frames made in the last iteration were used for the final computation of eye motion. The strip height and sampling rate used for the final strip analysis were 25 pixels and 540 Hz, respectively. Fifth, during off-line analysis, we could interpolate the cross-correlation maps around where the peak occurs to achieve subpixel resolution (one 10th of a pixel, 0.12 arcmin) in computing eye motion.

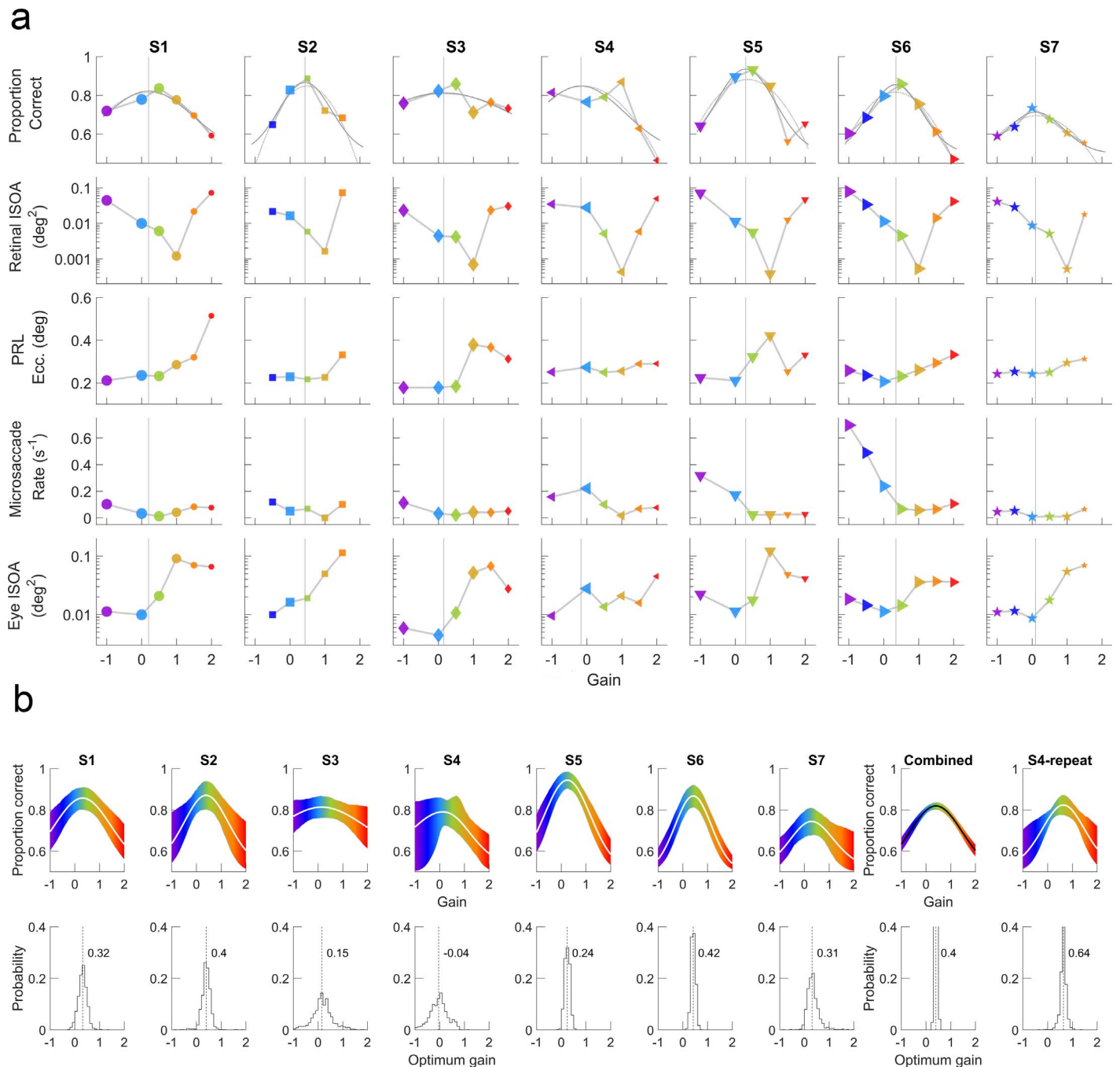


Figure 3. Individual results from Experiment 1. (a) Proportion correct and all mediators quantified in the present study, binned based on gain. (b) Bootstrapping tuning curve fits (top) and optimal gains (bottom) for each subject by using binary data (correct vs. incorrect). White lines represent the fits corresponding to median parameters. Shaded regions (top) represent 2.5%–97.5% percentiles of the bootstrapped distributions of fitted curves. For each panel, bootstrapping was done by resampling the individual trial data with replacement 1,000 times. Vertical dashed lines (bottom) represent the median optimal gains. The data from the second run of S4 were combined with the first run and analyzed together and are shown here in the rightmost panels. The vertical axes in bottom panels in Panel b are cropped to 0.4 for better visibility.

## Postprocessing

Following strip analysis of individual videos, the computed eye motion traces were subjected to several postprocessing steps. First, eye motion traces were

“rereferenced” to a larger reference frame created by retinal videos recorded in a separate session in which subjects were asked to fixate at a different position on the scanning raster. This essentially allowed us to capture images from different parts of the retina and

tile them on a larger (“global”) reference frame. Rereferencing was needed because each and every video had a slightly different (“local”) reference frame (because reference frames were created for each video separately), and hence, the absolute values of the eye motion would differ across trials. This step is required also for computing the absolute retinal position of the stimulus across all trials for a given subject. Rereferencing was performed by adding a constant shift to the previously computed eye motion traces, and the amount of shift was defined as the position of the local reference frame on the global one. After rereferencing, eye motions were then converted to visual degrees, low-pass filtered (passband and stopband frequencies of 80 and 120 Hz, respectively, with 65 dB attenuation in the stopband), and median filtered (with a window of nine samples,  $\sim 17$  ms) to reduce frame-rate artifacts (30-Hz noise and its harmonics). Filtered traces were used to compute retinal motion of the stimulus (defined as the difference between stimulus motion and eye motion). We quantified the amount of eye and retinal motion on a trial-by-trial basis as the 68% isoline area (Castet & Crossland, 2012; referred to as ISOA in the main text), which corresponds to the area of the 0.68 cumulative probability isoline (Figure 1d). This also roughly corresponds to the area covered by 68% percent of the motion samples. In the case of retinal motion, this metric quantifies the retinal area traversed by the stimulus in a given trial, and for eye motion, it represents the area of the raster (in world-centered coordinates) over which the eye moved. The distribution of the ratio of retinal ISOA and eye ISOA reveals how well the stimulus was delivered in different conditions (see Figure 1e and f). The theoretical ratio for a given gain is defined as

$$\frac{ISOA_{ret}}{ISOA_{eye}} = \text{sgn}(1 - g)(1 - g)^2,$$

where  $g$  represents gain and  $\text{sgn}(\cdot)$  represents the signum function, which was introduced to differentiate between gains that result in the same retinal motion magnitude but in opposite directions (Figures 1f and 3c).

We also computed the preferred retinal locus (PRL) of the stimulus for each trial. PRL was defined as the retinal location corresponding to the highest probability density of stimulus presence. The probability densities were computed by the “kernel density estimation via diffusion” method (Botev, Grotowski, & Kroese, 2010) with a slight modification. More specifically, the kernel bandwidth was set to one sixth of the standard deviation of the eye (or retinal) motion (as in Kwon, Nandy, & Tjan, 2013). The median PRL in the trials in which the gain was zero (i.e., natural viewing) was taken as the location of the fovea, and trial-to-trial PRL eccentricity was calculated with respect to this quantity. Finally, we identified micro-

saccades by using a median-based velocity threshold (Engbert & Kliegl, 2003). Eye-motion traces from all trials were visually inspected to ensure that micro-saccade detection was performed correctly.

The PRL estimated when gain is zero reflects the true PRL. Because PRL is mostly determined by the position of the stimulus at the beginning of a trial (e.g., when gain is one, the stimulus will stay at the start position), the estimated PRLs in other gain conditions do not necessarily reflect the preferences of the subjects. However, they demonstrate the idiosyncratic eye movements that govern the starting position of the stimulus. Nevertheless, to keep a consistent nomenclature, we used the term PRL across all conditions.

## Statistics

In order to test the tuning and no-tuning hypotheses, we fit performance with a flat line and a quadratic polynomial (as well as a Gaussian although polynomial and Gaussian fits produced almost identical results) and compared the adjusted  $R^2$  values as a metric of goodness of fit.

Due to foveal presentation of the stimuli, different gains led to different idiosyncratic oculomotor behaviors, which could not be controlled during the experiments. We quantified several covarying factors, such as retinal ISOA, eye ISOA, PRL eccentricity, and micro-saccade rate. The exact choice of covarying factors was driven by the need to account for main retinal and eye movement-related metrics. It is possible to estimate retinal image velocity, acceleration, or components of retinal image motion parallel or perpendicular to the orientation of the gratings. It is also possible to quantify these metrics in multiple ways, such as by their mean, standard deviation, minimum, maximum, or any combination of these together. However, because most of these metrics are strongly related to each other, adding different variants of them does not add much explanatory power. In addition, eye position traces extracted from retinal videos tend to have frame-rate artifacts, i.e., more power than normal at temporal frequencies around the frame rate of the videos. The frame-rate artifact gets amplified for velocity and acceleration due to differentiation, and more importantly, the severity of the effect interacts with different gain conditions (due to changes in oculomotor behavior). Position estimates are not influenced as much and indirectly capture the effect of their derivatives.

In order to determine how much of the effect of gain on performance is mediated by the aforementioned covarying factors, we performed a linear-mixed effects regression-based mediation analysis (MacKinnon, Fairchild, & Fritz, 2007). We followed the commonly used four-step approach suggested by Baron and



Kenny (1986). The aim of this analysis was to determine whether or not gain had a significant direct effect on discrimination performance even after taking into account the effects of mediators (i.e., significant covarying factors). Mediation analysis can be done in many ways depending on the causal relationship between the independent variable and mediators. When there are multiple mediators, say  $n$ , there are  $2^{(n!)}$  possible ways of decomposing total effect size. In our case,  $n = 4$ ; this yields 16,777,216 possibilities (Daniel, De Stavola, Cousens, & Vansteelandt, 2015). Because finding the best way of organizing mediators to account for data is beyond the scope of the present work, we chose a simple case in which all mediators were treated as independent factors that are directly modulated by only the independent variable, the gain, and they did not have interactions among each other nor with the gain. However, they were allowed to covary with subjects, and each had a fixed slope and a random intercept to account for individual differences in mediator values.

## Results

Using a TSLO (Sheehy et al., 2012), we presented seven human subjects with a high spatial frequency grating (12 c/deg) for 900 ms while imaging their retina and tracking their eye movements in real time (Figure 1a and b). We systematically manipulated the way retinal image motion and the actual FEM are related. The motion of the stimulus on the scanning raster was a function of the estimated eye motion times a *gain* factor (Figure 1d). A gain of zero means that the stimulus position remained fixed relative to the raster but slipped across the retina based on natural FEM. A gain of one means that the stimulus was stabilized on the retina, i.e., the retinal image motion due to FEM was completely cancelled out. A gain of 0.5 refers to partial stabilization, i.e., the stimulus moved only half as much as the eye motion. Assuming similar oculomotor behavior under different gains, a gain of  $-1$  doubles the retinal slip of the stimulus compared to that under natural viewing (i.e., gain = 0), and a gain of two results in the same retinal slip but in the opposite direction of what would occur under natural viewing. Off-line analyses of retinal videos for eye movement extraction revealed that eye tracking and stimulus delivery were performed with near-perfect accuracy ( $\sim 99\%$ ) for complete retinal stabilization (gain = 1). For gain conditions other than zero and one, there was some trial-to-trial variability in accuracy of stimulus delivery (Figure 1e and f). Nevertheless, each gain condition resulted in a statistically distinct distribution of effective gains centered at the desired gain (Figure 1e

and f). We measured subjects' ability to discriminate the direction of the grating's orientation from vertical under different gain conditions. If FEM are not tuned for fine discrimination at the fovea, then performance should not depend on gain (the null hypothesis, Figure 1c). On the other hand, if the retinal image motion due to FEM is tightly tuned for fine discrimination, performance should manifest a nonmonotonic relationship with gain, in which a particular value of gain results in the best (or worst) performance (Figure 1c). The tuning hypothesis can hold true in various ways. If FEM are optimal for fine discrimination at the fovea, then visual performance should peak at the gain of zero. Alternatively, retinal image motion might be the primary determinant of visual performance. If retinal image motion is always detrimental for seeing, visual performance should be highest at the gain of one. Retinal motion might also be beneficial regardless of the underlying FEM. If that is the case, the lowest discrimination performance should occur when the stimulus is fully stabilized on the retina.

We found that orientation-discrimination performance is tuned to gain (Figure 2a and b). Averaged across subjects, the peak performance occurred at a gain of 0.43 (95% confidence intervals: 0.12, 0.74), suggesting that partially reducing the effects of FEM is actually helpful in seeing fine spatial details. Results were similar when data from each subject were fitted separately: polynomial,  $t(6) = 3.165$ ,  $p = 0.019$ ; Gaussian,  $t(6) = 2.600$ ,  $p = 0.041$  (Figures 2b and 3). The exact choice of the tuning model (a quadratic polynomial or a Gaussian) did not matter: paired  $t$  test,  $t(12) = 3.165$ ,  $p = 0.019$ . Bootstrapping tuning curve fits to binary data (correct vs. incorrect) also revealed no optimality in six of the seven subjects (Figure 3). These results suggest that FEM are tuned but not optimal for fine discrimination at the fovea, at least within the range of parameters investigated here.

In order to check whether or not the tuning between performance and gain is limited only to fine discrimination tasks, we repeated the experiment with a spatial frequency (3 c/deg) at which the human visual system has the highest contrast sensitivity for static displays (Kelly, 1977). The hypothesis was that the retinal jitter due to FEM causes much lower modulations in RGCs with low spatial frequency stimuli; therefore, gain manipulations should result in minimal or no change in performance. We found no effect of gain on performance (Figure 4a) despite the retinal image motion varied over two log units across conditions (Figure 4d). Different performance trends with gain in Experiments 1 and 2 cannot be explained by differences in retinal motion, eye motion, microsaccade rate, or PRL eccentricity (Figure 2d through g vs. Figure 4d through g).

Next, we sought to determine what drives the strong dependency between performance and gain with high

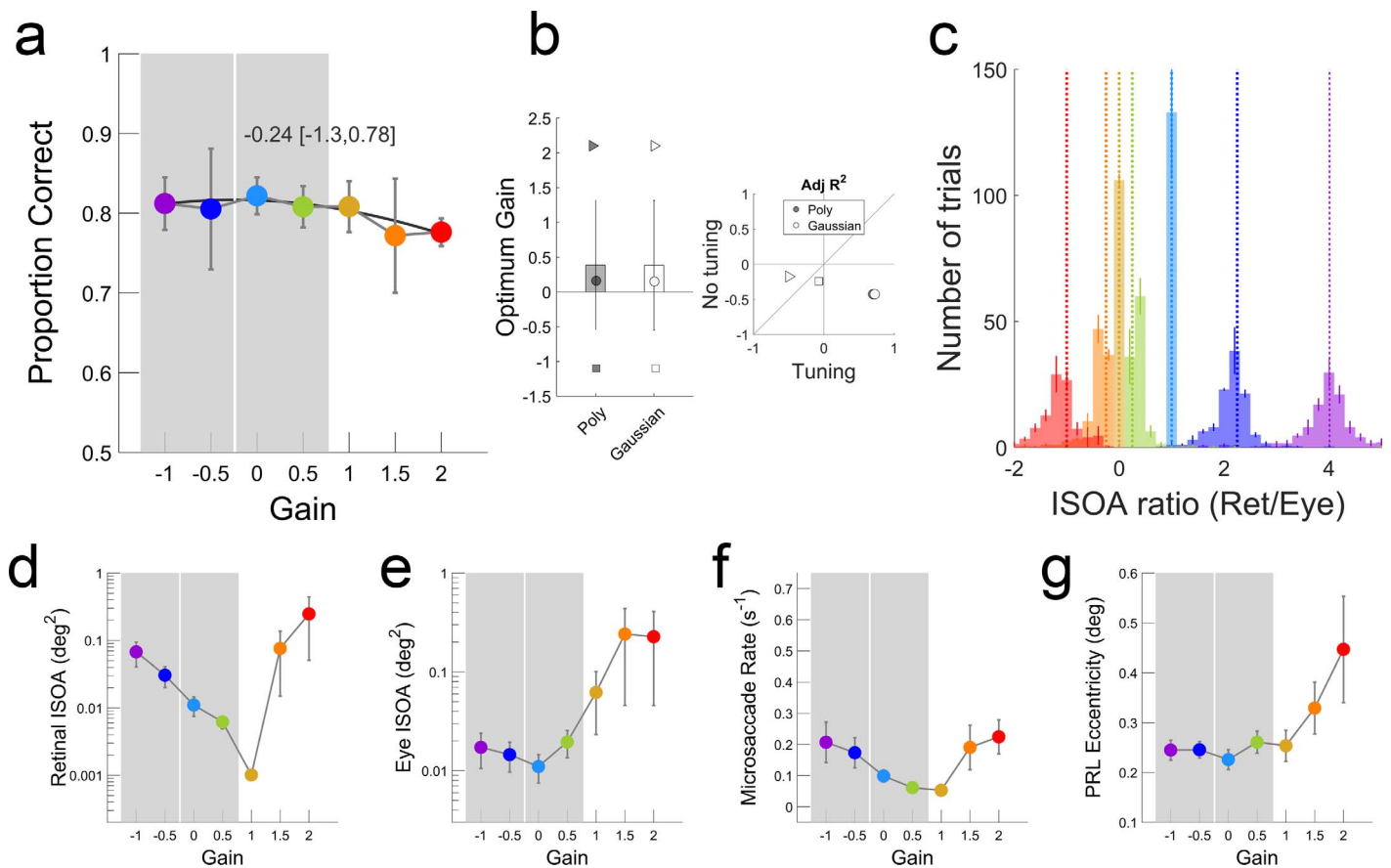


Figure 4. FEM are not tuned for coarse discrimination at the fovea. (a) Proportion correct as a function of gain in Experiment 2. (b) Average optimal gains based on individual tuning function fits along with individual optimal gains. (c) The distribution of retinal/eye motion ISOA in Experiment 2. (d, e) Retinal image motion and eye motion ISOA as a function of gain. (f, g) Microsaccade rates and PRL eccentricity across gains. Optimal gain and confidence intervals in Panel a are replotted in Panels d through g. Error bars represent  $\pm SEM$  ( $n = 7$ ). Conventions are as in Figure 2.

spatial frequency gratings. If gain manipulation only modulates the retinal image motion, then the answer would simply be retinal image motion, assuming no interference from extraretinal mechanisms. The approach taken in most retinal stabilization studies in the literature implicitly assumes that gain manipulation only results in changes in retinal image motion. In other words, retinal image motion is considered as the one and only mediator of performance. However, we found that gain modulates multiple mediators. We computed two-dimensional probability density of stimulus locations on the retina and eye positions on the raster (e.g., Figure 2d), and quantified, on a trial-by-trial basis, the extent of retinal image motion and eye motion by the ISOA containing roughly 68% of the retinal/eye motion traces (Figures 2d, e, and 3). As expected, the minimum retinal ISOA occurred when the stimulus was stabilized on the retina (i.e., gain = 1), but the pattern of changes in retinal ISOA as a function of gain revealed an asymmetric “V” shape around the gain of one (Figure 2d). This asymmetry can be explained by differences in oculomotor behavior of subjects across different gains.

More specifically, consistent with previous literature (Poletti, Listorti, & Rucci, 2010), subjects made smooth pursuit-like eye movements for gains of one and larger, which resulted in larger eye ISOAs (Figure 2e). This change in behavior occurs as soon as the retinal slip is no longer in a direction that is consistent with eye motion, in line with recent perceptual observations (Arathorn, Stevenson, Yang, Tiruveedhula, & Roorda, 2013). In addition, subjects made slightly more microsaccades for negative gains when retinal image motion was amplified (Figures 2f and 3). In addition, although each trial started with a fixation cross at the center of the raster, the PRL during grating presentation, defined here as the retinal location corresponding to peak probability density of retinal stimulus locations, also changed with gain (Figures 2c, f, g, and 3). To determine what really drives the relationship between gain and performance, one must take these mediators into account. In a regression-based mediation analysis following the most commonly used four-step approach (Baron & Kenny, 1986), we found that (a) gain has a significant effect on performance; (b) gain significantly

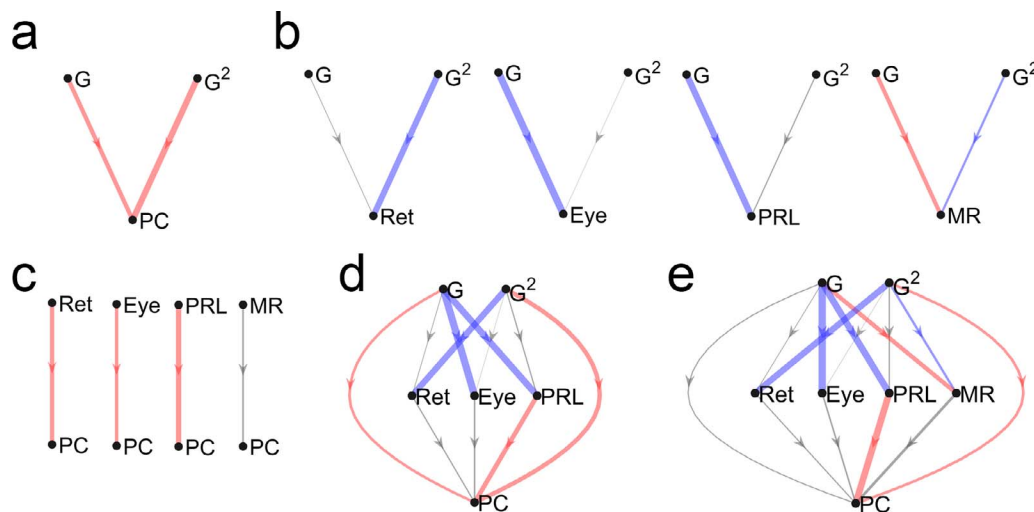


Figure 5. Teasing apart contributions of different mediators. (a) The first step in mediation analysis is to establish a significant relationship between gain ( $G$ ) and proportion correct ( $PC$ ). Because the tuning hypothesis predicts a quadratic relationship between  $G$  and  $PC$ , we included the  $G^2$  in our regression analyses. (b) Second, whether or not gain is a significant predictor of each covarying factor (Ret: retinal ISOA, Eye: eye ISOA, PRL: PRL eccentricity, MR: microsaccade rate) is established. (c) The third step tests separately for a significant effect of each mediator on performance. (d) Finally, gain and mediators with a significant correlation on performance are used to explain performance. Red and blue colors represent statistically significant negative and positive effects whereas gray lines represent insignificant relationships. The final model in Panel d shows that even when all significant mediators are taken into account, gain still has a significant effect on performance. (e) When all mediators are included, regardless of the outcome of Panel c, gain remains to be a significant factor. Thickness of each line represents the absolute value of the standardized effect size.

modulated all four mediators (retinal ISOA, eye ISOA, microsaccade rate, and PRL eccentricity); (c) all mediators individually, with the exception of microsaccade rate, are significant predictors of performance; and (d) gain remains a significant predictor of performance even when the effects of all significant mediators are taken into account (Figure 5). In order to determine whether or not mediators can account for the data as well as gain by itself, we performed a series of linear-mixed effects regression analyses (Figure 6). In terms of explained variance and log-likelihood, with which number of factors is not penalized, several purely mediator-based models could surpass the models based on gain only, suggesting that mediators identified here might fully account for how gain modulates performance. However, as the Bayes information criterion (BIC) differences show, none of the mediator-based models could outperform the simple model that is based only on gain. Through additional regression analyses and model comparisons using BIC, we confirmed that performance cannot be fully accounted for by mediators alone (Figure 6).

## Discussion

“Tuning” refers to a relationship between an independent variable and an outcome measure, with which a certain level of the former is more preferable

than others. Optimality in this context refers to achieving the best possible outcome in the face of several antagonist factors. Throughout the vast literature on FEM and visual perception, the word “optimal” has been used quite liberally in regard to spatiotemporal properties of FEM (Ahissar & Arieli, 2012; Cornsweet, 1956; Ditchburn, Fender, & Mayne, 1959; Gerrits & Vendrik, 1970; Kuang, Poletti, Victor, & Rucci, 2012; Martinez-Conde et al., 2004; Skavenski, Hansen, Steinman, & Winterson, 1979) although there has never been an explicit test for addressing it. Here, we tested whether visual performance in a fine orientation-discrimination task would show tuning as a function of the relationship between the retinal image motion and actual eye movements. We found strong tuning for a fine-detail discrimination task (Experiment 1) but not for a coarse discrimination task (Experiment 2). The absence of tuning in Experiment 2, despite up to a two log-unit change in retinal motion across conditions, suggests a very high tolerance for motion. Surprisingly, the optimal gain in Experiment 1 was obtained at a gain value between zero and one, suggesting that partially compensating for FEM can be beneficial.

Our results might seem inconsistent with previous reports in which complete retinal stabilization resulted in impaired discrimination performance (Ratnam et al., 2017; Rucci et al., 2007). A simple interpolation between the two extremes suggests a monotonic impairment in visual performance with better com-



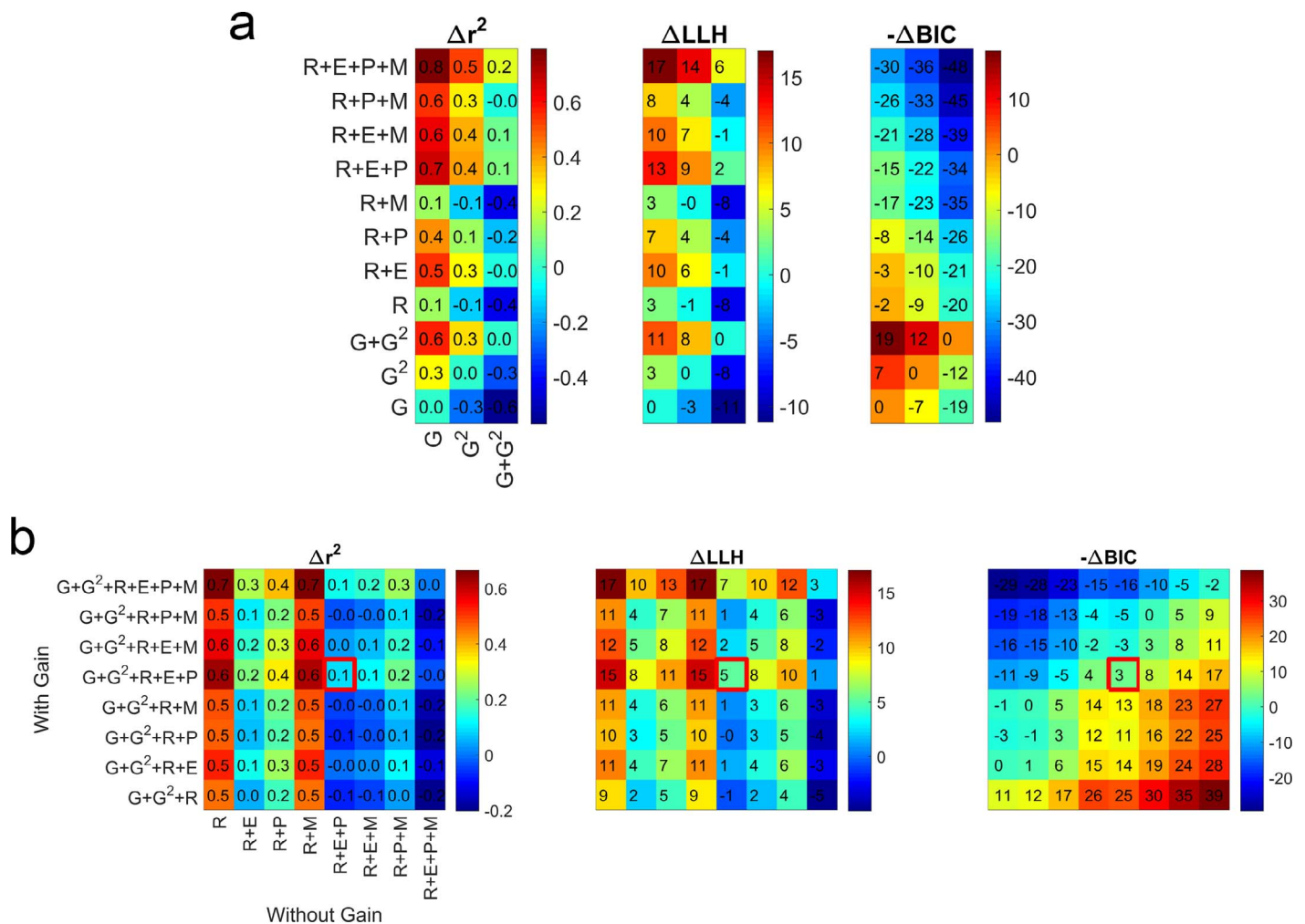


Figure 6. Contributions of gain and mediators in explaining variance. (a) Change in (left) explained variance, (middle) log-likelihood, and (right) BIC with addition of mediators. Note that the sign of  $\Delta\text{BIC}$  is flipped so that red color represents superiority of a model on the vertical axis with respect to another one on the horizontal axis. G: gain, R: retinal ISOA, E: eye ISOA, P: PRL eccentricity, M: microsaccade rate. (b) Can mediators fully account for the effects of gain? Here, we explicitly tested whether having gain in addition to mediators improve statistical models substantially. The right diagonal in each panel represents the exact contribution of the gain term. The red squares represent the final model in the mediation analysis ( $G + G^2 + R + E + P$ ; Figure 5d). In general, adding gain was helpful only when there were three or fewer mediators in the regression model.

pensation for FEM. This apparent inconsistency may not be real. First, it is technically possible to get impaired performance with complete stabilization *and* a nonzero optimal gain at the same time (which was the case for five out of seven subjects; Figure 3). Second, none of the existing studies explored the range of gains used here for discrimination tasks at the fovea. In addition, in previous studies, fading that resulted from retinal stabilization was quantified by threshold elevations, but the degree to which fading occurs depends on many variables, such as stimulus duration, size, contrast, eccentricity, equipment used, etc. (Coppola & Purves, 1996; Kelly, 1979a; Kelly, 1979b; Riggs et al., 1953; Riggs & Tulunay-Keesey, 1959). Early studies on retinal stabilization used small stimuli extending only a few arcmin and closely

surrounded by other visual cues coming from the apparatus (Ditchburn & Ginsborg, 1953; Riggs et al., 1953; Yarbus, 1967). More recent studies used foveally presented gratings extending several degrees of visual angle far from display boundaries (Poletti et al., 2013; Rucci et al., 2007) or parafoveally presented diffraction-limited stimuli covering only a few cones within a visible raster covering  $1^\circ$ – $1.3^\circ$  (Ratnam et al., 2017). The paradigm used here was somewhere in between; we presented through natural optics of the eye a grating that covers the fovea and is situated within a visible raster covering  $10^\circ$ . Therefore, it is possible to make qualitative comparisons across the aforementioned studies. However, it is not feasible to extrapolate previous studies to the conditions investigated here.



Our results are highly consistent with recent theoretical work that has been successfully used to account for performance impairment due to retinal stabilization (Anderson, Olshausen, Ratnam, & Roorda, 2017; Burak et al., 2010; Pitkow, Sompolinsky, & Meister, 2007). According to this framework, there are two distinct mechanisms that work in tandem: one for estimating FEM from RGC responses across the retina, which negates the need for an extraretinal mechanism to properly decode spatial information, and another one for making an *optimal* inference about the spatial layout of the stimuli. The presence of a global motion compensation mechanism for FEM was demonstrated by a striking visual illusion (Murakami & Cavanagh, 1998). Surprisingly, when receptive field size and density across the retina and the statistics of FEM under normal viewing conditions are factored in, this model predicts that normal human FEM are not optimal for high-acuity tasks (Burak et al., 2010; Pitkow et al., 2007). This theory also predicts that larger stimulus sizes and peripheral cues would improve discrimination at the fovea because estimating FEM would be easier and more accurate in these conditions. It is possible that the absence of optimality might have arisen because the scanning raster was always visible in the present work. Although several lines of evidence against this prediction have been presented (Wehrhahn, 2011), they turned out to be lacking technical precision and proper controls to directly test this prediction (Burak, Rokni, Meister, & Sompolinsky, 2011).

## Previous work on FEM using retinal image stabilization

Although there have been many studies on the effects of FEM on various tasks since the 1950s, many of them are not directly comparable to ours. This is in part due to the significant improvements in eye tracking and stimulus presentation techniques and stimulus types and properties over the years. Nevertheless, in order to inform the reader of these highly informative studies, we briefly review some of them here.

Since the early work by Ditchburn and Ginsborg (1952), Riggs et al. (1953), and Yarbus (1967), in which image fading with stabilization was demonstrated, the effects of FEM on spatial (Gilbert & Fender, 1969; Kelly, 1979a; Tulunay-Keeseey & Bennis, 1979; Tulunay-Keeseey & Jones, 1976; Watanabe, Mori, Nagata, & Hiwatashi, 1968) and spatiotemporal contrast sensitivity (Kelly, 1977; Kelly, 1979b; Kelly, 1981b), chromatic contrast sensitivity (Kelly, 1981a), detection of colored light (Ditchburn & Foley-Fisher, 1979), Vernier acuity (Tulunay-Keeseey, 1960), edge, line, or overall form detection (Gerrits & Vendrik, 1970b; Gerrits & Vendrik, 1974; Tulunay-Keeseey, 1960a),

orientation discrimination (Rucci et al., 2007; Tulunay-Keeseey, 1960), and retinal eccentricity (Gerrits, 1978) have been documented. Interestingly, there were large differences across studies in how much FEM affect perception. For instance, although similar contrast sensitivity functions were obtained under normal viewing conditions, the contrast threshold elevation under stabilization ranged from zero (no effect at all) up to >10 times (1.0 log unit) across studies (Gilbert & Fender, 1969; Kelly, 1979a; Tulunay-Keeseey & Bennis, 1979; Tulunay-Keeseey & Jones, 1976; Watanabe et al., 1968). The differences in precision of retinal image stabilization (Gerrits, 1978) as well as the stimulus duration (Tulunay-Keeseey & Jones, 1976; Tulunay-Keeseey & Jones, 1980) have been identified as the primary determinants of these differences, and more precise stabilization and longer stimulus duration are associated with larger threshold elevation.

More interestingly, there is also a glaring inconsistency between these early studies and more recent work: the spatial frequency dependency of contrast elevation with stabilization. Tulunay-Keeseey and Jones (1976) found that threshold elevation due to stabilization was negligible up to a 1,000-ms stimulus duration whereas a uniform reduction in sensitivity across all spatial frequencies were found when observers were allowed to view the stimulus indefinitely. Tulunay-Keeseey and Bennis (1979) reported maximum threshold elevations at middle spatial frequencies (around 3 c/deg, 0.3 log units elevation) only when the stimuli were temporally ramped up and down. For step-onset stimuli, stabilization did not have any effect across all spatial frequencies. However, when observers were instructed to wait for image fading before searching for contrast threshold, a completely different pattern of threshold elevations emerged: Image stabilization reduced contrast sensitivity for low spatial frequencies (e.g., 0.5 c/deg) by more than 10 times whereas sensitivity to high spatial frequencies was less affected (e.g., approximately two times reduction at 10 c/deg). Using method of adjustment, Gilbert and Fender (1969) and Kelly (1979a) also showed larger contrast threshold elevations for low spatial frequencies. However, note that Kelly (1979a) also found larger elevation at high spatial frequencies for young (~30 years old) subjects. He hypothesized that this discrepancy might be due to the differences in accommodative abilities with age because accommodation affects mostly high spatial frequencies and it is virtually absent after a certain age. Watanabe et al. (1968) reported a constant drop in sensitivity within the range of spatial frequencies that they investigated (0.1–5 c/deg). More recent work by Rucci et al. (2007) reported very little or no effect of stabilization for low- and medium-range spatial frequencies (<4 c/deg) whereas they found increasing contrast threshold elevations with increasing

spatial frequencies beyond 4 c/deg. Admittedly, the size, orientation, and duration of the gratings; adapting luminance; task; observer characteristics, etc., all contribute to contrast thresholds even during fixation. Although the inconsistencies across studies might be partially accounted for by the differences in the aforementioned parameters, it is still very difficult to make quantitative comparisons across studies considering additional differences in the accuracy and precision of the stabilization methods.

There are few studies in which retinal image motion was systematically controlled (Ditchburn et al., 1959; Gerrits & Vendrik, 1974; Riggs & Tulunay-Keesey, 1959). Ditchburn et al. (1959) imposed controlled oscillatory retinal motions after annulling the naturally occurring FEM and measured the percentage of time during which a vertical line was perceived. For both simulated drifts (by large-amplitude, low-frequency oscillations) and simulated microsaccades (by means of 1-ms jumps) imposed on the stabilized image, they found a nonmonotonic trend for visibility of the test object. For simulated drifts, the maximum visibility occurred at an oscillation amplitude close to the maximum (not the median) amplitude of the naturally occurring ocular drift. For simulated microsaccades, visibility was similar across all simulated amplitudes (5–25 arcmin). Note that although these findings suggest that increasing the drift amplitude might be better for visibility, this and other early studies on different types of FEM underestimated ocular drifts by 1–20 arcmin/s (Ditchburn & Ginsborg, 1953; Ratliff & Riggs, 1950) versus median  $\sim 50$  arcmin/s (Kuang et al., 2012). In addition, because drifts are almost never pure sinusoidal oscillations, the controlled manipulation in Ditchburn et al. (1959) did not reflect different “gains” as in the present study.

In order to determine the characteristics of eye movements that are needed to preserve normal perception, Gerrits and Vendrik (1974) performed experiments in which a  $4^\circ \times 4^\circ$  square was moved on the retina in a controlled way. After stabilizing the image of the square on the retina by a suction cup, they modulated the motion of the square in various ways and asked observers to rate their percepts in five categories (from “only contours” up to “homogeneous square”). Multiple types of motion were investigated: sinusoidal or triangular oscillations of varying frequencies and amplitudes, in- or out-of-phase motion for horizontal and vertical dimensions, and finally Gaussian noise and binary noise with varying strengths. They found that high-frequency ( $>2$  Hz), low-amplitude oscillations were not effective in preserving the perception of homogeneous square whereas low-frequency ( $<0.2$  Hz), large-amplitude (e.g.,  $3^\circ$ ) oscillations were most effective in preserving normal vision. Moreover, they found that the closest case to

natural viewing was obtained when the retinal motion was constructed by a Gaussian + binary noise with which the former simulates drifts and the latter represents microsaccades. They also reported that binary noise only was ineffective in keeping the normal vision. They concluded that the irregular and continuous nature of drifts enables us to preserve normal vision.

The manipulations of retinal image motion that are closest to the conditions in the present study were performed by Riggs and Tulunay-Keesey (1959), with which the “gain” of the retinal image motion could be controlled by varying the distance between a pair of mirrors in an optical setup. The outcome measure was again the percentage of time during which a test object (a disk consisting of two semicircles differing in luminance) was seen. The range of gain values was 0.74 to 2.25. Within these values, they found a nonmonotonic relationship between visibility and gain with which the worst visibility occurred at a gain of one. Unfortunately, we cannot extrapolate from this data to gain values below 0.74 to compare with our findings.

## Covarying factors

We have identified several mediator factors that could explain a significant portion of the variability in the data. Note that the presence of these mediators is not due to the equipment used or stimulus parameters, but reflects the inevitable consequence of foveal presentation of the stimuli. None of these mediators have been reported quantitatively or used to account for data in the previous literature about the roles of FEM. Parafoveal (or peripheral) presentation of stabilized stimuli may not activate all of the aforementioned mediators (e.g., eye ISOA); however, non-foveal presentation of stimuli would defeat the purpose of this study because one cannot make strong inferences about foveal viewing with peripherally presented stimuli. Alternatively, an experiment in which stimulus moves in an incongruent manner to avoid chasing can be performed; however, it is unclear whether or not small amplitudes of stimulus motion would still lead to pursuit-like eye movements. The way we chose to address what factors underlie the tuning between performance and gain reported here was to perform a mediation analysis (MacKinnon et al., 2007). This analysis showed that, even when retinal motion, eye motion, PRL eccentricity, and microsaccade rate were factored in, gain still had a significant direct effect on performance. This finding suggests that (a) there are additional mediators not considered here or (b) “postretinal” factors, such as changes in attentional engagement in the task depending on gain value, might be at play.

In order to assign extraretinal factors a role for perception during FEM, one needs to factor out all possible retinal factors, such as retinal ISOA, velocity, acceleration, PRL eccentricity, initial retinal position of the stimuli, etc. Obviously, these factors are not independent from each other, limiting the use of mediation analysis described here. Admittedly, the optimal gain might also be affected by these mediators. A way to compensate for their effects for the purpose of estimating optimal gain might be normalizing performance by each mediator and then testing for tuning. However, this exacerbates the problem because (a) whether a covarying factor is a positive mediator (reducing the effect) or a negative one (increasing the effect) is not known a priori; (b) the relative contribution of each mediator is different but normalization assumes equal contribution; and (c) each mediator has a different scale of change across conditions, which could result in numerical instabilities and prevent accurate determination of the optimal gain. Point c can be addressed by log-transforming some mediators (e.g., retinal ISOA) and/or standardizing them, and point a can be addressed by using the outcome of a mediation analysis to guide the normalization process, but point b cannot be readily addressed. On the other hand, because visual performance comes about via mediators, there may not be a need for normalizing performance before computing optimal gains. From this perspective, they are not just artifacts to be removed, but the actual underlying factors of visual function. The logic is that whatever the exact value of optimal gain is, visual performance results from an interplay between various mediators, and it may not be possible to uniformly sample the multidimensional space defined by multiple mediators. A case in point, it seems that foveal presentation of a stimulus almost always leads to smooth pursuit-like oculomotor behaviors when the retinal projection of it is stabilized (Poletti et al., 2010).

## Microsaccades

The human retina is nonhomogeneous even within the fovea. Therefore, making microsaccades to redirect gaze to enjoy the highest acuity part of the retina is a reasonable strategy (Cornsweet, 1956; Poletti et al., 2013). Microsaccades are not always initiated voluntarily, however, and recent studies claimed that they often occur after a period of low retinal slip and are executed to avoid fading (Engbert & Mergenthaler, 2006). Their occurrence seems to be coupled to heartbeat as well (Ohl, Wohltat, Kliegl, Pollatos, & Engbert, 2016). Based on the finding that microsaccades cause widespread activity across the visual system and help temporally synchronize neural activity, some researchers supported the view that microsac-

cades, among other FEM, contribute most to visual function (Martinez-Conde et al., 2013; Masquelier, Portelli, & Kornprobst, 2016; McCamy et al., 2012). In addition, a review of old and new literature on microsaccades led some researchers to conclude that microsaccades do not serve a useful purpose (Colleijn & Kowler, 2008). Nevertheless, in order to address these hypotheses, we performed a series of analyses on microsaccades made by all observers (Figure 7). Because the rate of microsaccades was rather low in our experiments, we combined the data across observers for the following analyses. The low rate of microsaccades itself, especially when retinal image motion was minimized, is evidence against a primary role for microsaccades for visual processing. In response to partial or complete retinal stabilization, for instance, subjects made larger drifts rather than larger or more frequent microsaccades. Moreover, we found evidence for gaze redirection; most microsaccades were made to bring the retinal projection of the stimuli closer to the PRL (Figure 7b), consistent with Poletti et al. (2013); H. Ko, Poletti, and Rucci (2010); and Chen and Hafed (2013). Chen and Hafed, in fact, specifically investigated the premicrosaccadic eye velocity traces and showed no decrease in eye velocity prior to microsaccade onset. These authors further showed that prior reports of low-retinal slip just before microsaccades (e.g., Engbert & Mergenthaler, 2006) suffered from artifacts of video-based trackers. They recorded eye movements of monkeys with a video-based tracker and with a search-coil technique simultaneously and found that radial eye velocity dropped significantly only in the data from the video-based eye tracker whereas the data from the search coils did not show such a reduction. A more detailed study, preferably powered by retinal imaging, to decisively determine the main role of microsaccades (see, for review, Rolfs, 2009), is necessary.

## Ocular drifts

There are several other facts to be considered when functional roles of drifts and microsaccades are to be determined. First, RGCs are most responsive to light transients, and the time constant of their responses can vary from 30 to 100 ms (O'Brien, Isayama, Richardson, & Berson, 2002). Second, although the initial burst activity of RGCs in response to a light transient is highly precise, prolonged presentation breaks this temporal synchrony, and the tonic neural activity demonstrates quite a bit of variability (Berry, Warland, & Meister, 1997; Reich, Victor, Knight, Ozaki, & Kaplan, 1997). Encoding spatial information using a rate code with a few spikes necessitates the accumulation of information over time to improve the signal-to-



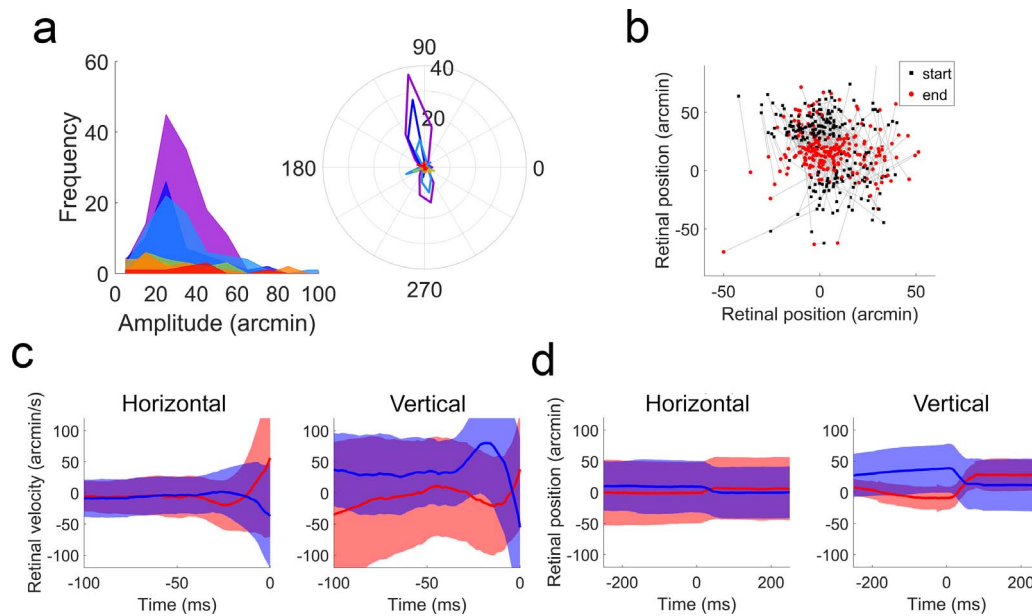


Figure 7. Analyses of microsaccades. (a) Amplitude and direction distribution of microsaccades combined across seven subjects in Experiment 1. (b) The retinal position of the stimuli at the start (black squares) and end (red circles) of microsaccades. Clearly, the primary role of microsaccades was redirecting gaze to compensate for nonhomogeneous vision. (c) Retinal image velocity just before microsaccades. The blue and red lines represent downward and upward microsaccades (within  $\pm 45^\circ$  from vertical was considered as upward). (d) Retinal position of the stimuli across microsaccades. In Panels c and d, the panels on the left and right represent data from horizontal and vertical component of the eye movements, respectively.

noise ratio. The presence of FEM makes encoding of spatial information via rate coding even less reliable by further increasing variability in spiking activity. Third, FEM create retinal motion signals that are well beyond motion detection thresholds but not perceived.

From an evolutionary standpoint, it is unclear which of the facts listed so far was the root cause for the others. For instance, whether RGCs prefer light transients and do not respond as strongly after prolonged presentation due to FEM or FEM exist due to the temporal characteristics of RGC responses is a hard problem to address. In addition, a recent modeling work demonstrated potential alternatives to spatial encoding via rate coding, with which FEM do not pose problems to be solved by the visual system, but instead they are part of the solution to efficient information encoding (Ahissar & Arieli, 2012). This also renders the mystery of how the visual system differentiates motion due to FEM from those of external objects a nonissue because if we actually see via FEM, why correct for them? In fact, drifts transform the spectral content of retinal stimulation into spatiotemporal frequencies to which the early visual system is most sensitive (Kuang et al., 2012), but it is unclear whether this is an epiphenomenon or a result targeted by an active and/or adaptive process. However, the current implementation of this model relies on weak assumptions, one of which is that drifts are cyclic (sinusoidal) motions (to drive phase-locking

mechanism) within time courses that reflect average fixation duration ( $\sim 300$  ms). Except in very few instances, we did not observe such patterns (Figure 8).

## Abnormal FEM

Some visual/cortical impairments (e.g., amblyopia, central vision loss) result in “abnormal” FEM (Chung, Kumar, Li, & Levi, 2015; Kumar & Chung, 2014). In a computer vision system with limited spatial resolution or blurry optics, it is theoretically possible to achieve “super-resolution” or deblurring by moving a sensor array. Therefore, we think that to classify FEM as abnormal, one needs to consider several factors, such as the amount of blur, receptive field sizes, and contrast sensitivity at the PRL. Otherwise, a genuine strategy of a perfectly normal oculomotor system might be misinterpreted as an artifact. In the case of central vision loss, the use of peripheral PRL leads to changes in all these factors, and it is quite possible that apparently abnormal FEM in these patients might be a way to compensate for these changes. In fact, recent studies on the effects of retinal image motion in normal peripheral vision reported improvements in reading and discrimination performance with increased motion (Patrick, Roach, & McGraw, 2017; L. M. Watson et al., 2012).



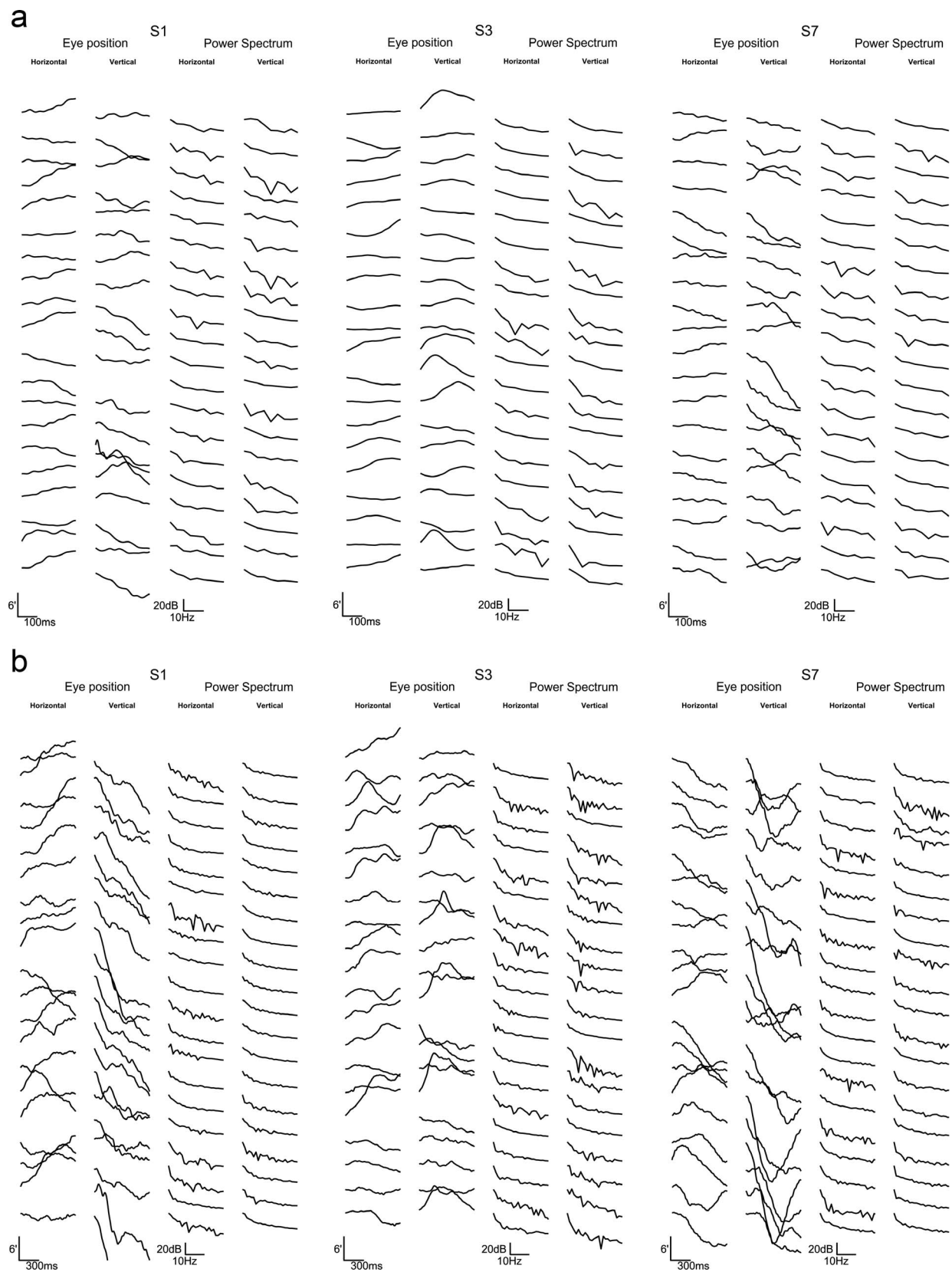


Figure 8. Ocular drifts under normal viewing conditions. To qualitatively test the assumption that drifts are cyclic motions within the time scale of typical fixation ( $\sim 300$  ms), we randomly sampled 20 eye position traces from three subjects in the zero-gain condition and computed the power spectra of both the horizontal and vertical components in (a) 300-ms and (b)  $\sim 900$ -ms time windows. If

→

←

indeed, drifts show three to five cycles per  $\sim 300$  ms; this should be visible as clear peaks in the power spectra and disappear when power spectra are computed over a longer time scale. However, except for a few cases, we did not encounter such motions. For some subjects, much lower-frequency fluctuations were visible, but these fluctuations are too slow to be of any use for fast and efficient temporal encoding. For this particular figure, eye position traces were further filtered with a low-pass filter with a cutoff frequency of 25 Hz. Power spectra are shown with a linear frequency axis with limits from 3 to 30 Hz.

## Limitations and future directions

The statistics of FEM may change when a subject's head is restrained compared to head-free viewing (Poletti, Aytekin, & Rucci, 2015). The amplitudes of FEM increase under free viewing. Measurements using a Dual-Purkinje tracker showed that drifts from the two eyes show minimal correlation under head-fixed conditions, and they become mostly conjugate under head-free conditions. However, retinal imaging via a binocular TSLO revealed almost complete conjugacy under head-fixed conditions (Stevenson, Sheehy, & Roorda, 2016). Nonetheless, the conditions reported here may demonstrate a special case of oculomotor control, which need not be optimized because outside the laboratory, we always view the environment with a freely moving body and head. In addition, because the stimulus presentation was monocular in the present study, it remains to be seen whether similar tuning functions would be obtained with binocular presentation. It may be that binocular viewing increases the tolerance of the visual system to retinal image motion due to FEM even for high spatial frequencies due to redundancy from the second eye. In addition, as mentioned before, varying retinal image motion while keeping the eye motion unaffected remains a challenge. Finally, the different patterns of results in the two experiments reported here also suggest that the relationship between FEM and spatial frequency might be a continuum from no tuning to optimal tuning with increasing spatiotemporal frequency. Future endeavors along these lines will require denser sampling of the frequency space as well as accurate eye tracking combined with fast stimulus delivery to both eyes.

**Keywords:** *retinal imaging, microsaccades, ocular drifts, fixational eye movements, scanning laser ophthalmoscopy*

## Acknowledgments

We would like to thank Jonathan Patrick, Arun Kumar Krishnan, Haluk Ogmen for his comments on the manuscript, and Harold Bedell for many discussions of our results. This study was supported by grants R01-EY012810 (STLC), R01-EY017707

(STLC), R01-EY023591 (AR), P30-EY003176 (core grant) from the National Institutes of Health. MNA and STLC conceived the idea and designed the experiments. STLC provided resources (the TSLO system, subjects and seeking IRB approval). MNA and STLC performed the experiments. MNA analyzed the data and wrote the manuscript. CKS, PT, and AR provided technical support for the TSLO system (hardware, electronics, and software), and CKS, PT, AR, and STLC reviewed the manuscript.

**Commercial relationships:** MNA and STLC have no financial interest. AR, CKS, and PT hold a patent for the design of the TSLO and have financial interest in C.Light Technologies.

**Corresponding author:** Mehmet N. Ağaoğlu.

Email: mnagaoglu@gmail.com.

Address: School of Optometry, University of California, Berkeley, Berkeley, CA, USA.

## References

- Ahissar, E., & Arieli, A. (2012). Seeing via miniature eye movements: A dynamic hypothesis for vision. *Frontiers in Computational Neuroscience*, 6:89.
- Anderson, A. G., Olshausen, B. A., Ratnam, K., & Roorda, A. (2017). A neural model of high-acuity vision in the presence of fixational eye movements. In *Conference Record - Asilomar Conference on Signals, Systems and Computers* (pp. 588–592). Pacific Grove, CA: IEEE.
- Arathorn, D. W., Stevenson, S., Yang, Q., Tiruveedhula, P., & Roorda, A. (2013). How the unstable eye sees a stable and moving world. *Journal of Vision*, 13(10):22, 1–19, <https://doi.org/10.1167/13.10.22>. [PubMed] [Article]
- Arathorn, D. W., Yang, Q., Vogel, C. R., Zhang, Y., Tiruveedhula, P., & Roorda, A. (2007). Retinally stabilized cone-targeted stimulus delivery. *Optics Express*, 15(21), 13731–13744.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical

- considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.
- Benardete, E. A., & Kaplan, E. (1997). The receptive field of the primate P retinal ganglion cell, I: Linear dynamics. *Visual Neuroscience*, 14(1997), 169–185.
- Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences, USA*, 94(10), 5411–5416.
- Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5), 2916–2957.
- Burak, Y., Rokni, U., Meister, M., & Sompolinsky, H. (2010). Bayesian model of dynamic image stabilization in the visual system. *Proceedings of the National Academy of Sciences, USA*, 107(45), 19525–19530.
- Burak, Y., Rokni, U., Meister, M., & Sompolinsky, H. (2011). Reply to Wehrhahn: Experimental requirements for testing the role of peripheral cues in dynamic image stabilization. *Proceedings of the National Academy of Sciences, USA*, 108(10), E36.
- Castet, E., & Crossland, M. (2012). Quantifying eye stability during a fixation task: A review of definitions and methods. *Seeing and Perceiving*, 25(5), 449–469.
- Chen, C. Y., & Hafed, Z. M. (2013). Postmicrosaccadic enhancement of slow eye movements. *The Journal of Neuroscience*, 33(12), 5375–5386.
- Chung, S. T. L., Kumar, G., Li, R. W., & Levi, D. M. (2015). Characteristics of fixational eye movements in amblyopia: Limitations on fixation stability and acuity? *Vision Research*, 114, 87–99.
- Collewijn, H., & Kowler, E. (2008). The significance of microsaccades for vision and oculomotor control. *Journal of Vision*, 8(14):20, 1–21, <https://doi.org/10.1167/8.14.20>. [PubMed] [Article]
- Coppola, D., & Purves, D. (1996). The extraordinarily rapid disappearance of entopic images. *Proceedings of the National Academy of Sciences, USA*, 93(15), 8001–8004.
- Cornsweet, T. N. (1956). Determination of the stimuli for involuntary drifts and saccadic eye movements. *Journal of the Optical Society of America*, 46(11), 987–988.
- Costela, F. M., McCamy, M. B., Macknik, S. L., Otero-Millan, J., & Martinez-Conde, S. (2013). Microsaccades restore the visibility of minute foveal targets. *PeerJ*, 1, e119.
- Curcio, C., Sloan, K., Kalina, R., & Hendrickson, A. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, 4(292), 497–523.
- Dacey, D. M., & Petersen, M. R. (1992). Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences, USA*, 89(20), 9666–9670.
- Daniel, R. M., De Stavola, B. L., Cousens, S. N., & Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics*, 71(1), 1–14.
- Ditchburn, R. W., Fender, D. H., & Mayne, S. (1959). Vision with controlled movements of the retinal image. *Journal of Physiology*, 145, 98–107.
- Ditchburn, R. W., & Foley-Fisher, J. A. (1979). The effect of retinal image motion on vision in coloured light. *Vision Research*, 19(11), 1223–1227.
- Ditchburn, R. W., & Ginsborg, B. L. (1952, July 5). Vision with a stabilized retinal image. *Nature*, 170, 36–37.
- Ditchburn, R. W., & Ginsborg, B. L. (1953). Involuntary eye movements during fixation. *Journal of Physiology*, 119(1940), 1–17.
- Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, 43(9), 1035–1045.
- Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences, USA*, 103(18): 7192–7.
- Engbert, R., Mergenthaler, K., Sinn, P., & Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences, USA*, 108(39), E765–E770.
- Gerrits, H. J. M. (1978). Differences in foveal and peripheral effects observed in stabilized vision. *Experimental Brain Research*, 32, 225–244.
- Gerrits, H. J. M., & Vendrik, A. J. H. (1970). Artificial movements of a stabilized image. *Vision Research*, 10(12), 1443–1456.
- Gerrits, H. J. M., & Vendrik, A. J. H. (1974). The influence of stimulus movements on perception in parafoveal stabilized vision. *Vision Research*, 14(2), 175–180.
- Gilbert, D. S., & Fender, D. H. (1969). Contrast thresholds measured with stabilized and non-stabilized sine-wave gratings. *Optica Acta*, 16(2), 191–204.
- Kaplan, E., & Benardete, E. (1999). The dynamics of primate M retinal ganglion cells. *Visual Neuroscience*, 16, 355–368.
- Kelly, D. H. (1977). Visual contrast sensitivity. *Optica Acta: International Journal of Optics*, 24(2), 107–129.

- Kelly, D. H. (1979a). Motion and vision. I. Stabilized images of stationary gratings. *Journal of the Optical Society of America*, 69(9), 1266–1274.
- Kelly, D. H. (1979b). Motion and vision. II. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10), 1340–1349.
- Kelly, D. H. (1981a, December 11). Disappearance of stabilized chromatic gratings. *Science*, 214(4526), 1257–1258.
- Kelly, D. H. (1981b). Nonlinear visual responses to flickering sinusoidal gratings. *Journal of the Optical Society of America*, 71(9), 1051–1055.
- Ko, H., Poletti, M., & Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature Neuroscience*, 13(12), 1549–1553.
- Ko, H. K., Snodderly, D. M., & Poletti, M. (2016). Eye movements between saccades: Measuring ocular drift and tremor. *Vision Research*, 122, 93–104.
- Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51(13), 1457–1483.
- Kuang, X., Poletti, M., Victor, J. D., & Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology*, 22(6), 510–514.
- Kumar, G., & Chung, S. T. L. (2014). Characteristics of fixational eye movements in people with macular disease. *Investigative Ophthalmology & Visual Science*, 55(8), 5125–5133.
- Kwon, M., Nandy, A. S., & Tjan, B. S. (2013). Rapid and persistent adaptability of human oculomotor control in response to simulated central vision loss. *Current Biology*, 23(17), 1663–1669.
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58(1), 593–614.
- Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews. Neuroscience*, 5(3), 229–240.
- Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: Towards a unified theory of saccadic function. *Nature Review. Neuroscience*, 14(2), 83–96.
- Masquelier, T., Portelli, G., & Kornprobst, P. (2016). Microsaccades enable efficient synchrony-based coding in the retina: A simulation study. *Scientific Reports*, 6:24086.
- McCamy, M. B., Otero-Millan, J., Macknik, S. L., Yang, Y., Troncoso, X. G., Baer, S. M., ... Martinez-Conde, S. (2012). Microsaccadic efficacy and contribution to foveal and peripheral vision. *Journal of Neuroscience*, 32(27), 9194–9204.
- Mulligan, J. (1998). Recovery of motion parameters from distortions in scanned images. In *NASA Conference Publication* (pp. 281–292). NASA.
- Murakami, I., & Cavanagh, P. (1998, October 22). A jitter after-effect reveals motion-based stabilization of vision. *Nature*, 395(6704), 798–801.
- O'Brien, B. J., Isayama, T., Richardson, R., & Berson, D. M. (2002). Intrinsic physiological properties of cat retinal ganglion cells. *The Journal of Physiology*, 538(Pt 3), 787–802.
- Ohl, S., Wohltat, C., Kliegl, R., Pollatos, O., & Engbert, R. (2016). Microsaccades are coupled to heartbeat. *Journal of Neuroscience*, 36(4), 1237–1241.
- Patrick, J. A., Roach, N. W., & McGraw, P. V. (2017). Motion-based super-resolution in the peripheral visual field. *Journal of Vision*, 17(9):15, 1–10, <https://doi.org/10.1167/17.9.15>. [PubMed] [Article]
- Pitkow, X., Sompolinsky, H., & Meister, M. (2007). A neural computation for visual acuity in the presence of eye movements. *PLoS Biology*, 5(12), 2898–2911.
- Poletti, M., Aytakin, M., & Rucci, M. (2015). Head-eye coordination at a microscopic scale. *Current Biology*, 25(24), 3253–3259.
- Poletti, M., Listorti, C., & Rucci, M. (2010). Stability of the visual world during eye drift. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(33), 11143–11150.
- Poletti, M., Listorti, C., & Rucci, M. (2013). Microscopic eye movements compensate for nonhomogeneous vision within the fovea. *Current Biology*, 23(17), 1691–1695.
- Ratliff, F., & Riggs, L. (1950). Involuntary motions of the eye during monocular fixation. *Journal of Experimental Psychology*, 40(6), 687–701.
- Ratnam, K., Harmening, W. M., & Roorda, A. (2017). Benefits of retinal image motion at the limits of spatial vision. *Journal of Vision*, 17(1):30, 1–11, <https://doi.org/10.1167/17.1.30>. [PubMed] [Article]
- Reich, D. S., Victor, J. D., Knight, B. W., Ozaki, T., & Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. *Journal of Neurophysiology*, 77(5), 2836–2841.
- Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *Journal of the Optical Society of America*, 43(6), 495–501.
- Riggs, L. A., & Tulunay-Keesey, Ü. (1959). Visual effects of varying the extent of compensation for eye movements. *Journal of the Optical Society of America*, 49(8), 741–745.



- Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research*, 49(20), 2415–2441.
- Rucci, M., Iovin, R., Poletti, M., & Santini, F. (2007, June 14). Miniature eye movements enhance fine spatial detail. *Nature*, 447(7146), 852–855.
- Rucci, M., & Poletti, M. (2015). Control and functions of fixational eye movements. *Annual Review of Vision Science*, 1(1), 499–518, <https://doi.org/10.1146/annurev-vision-082114-035742>.
- Sheehy, C. K., Yang, Q., Arathorn, D. W., Tiruveedhula, P., de Boer, J. F., & Roorda, A. (2012). High-speed, image-based eye tracking with a scanning laser ophthalmoscope. *Biomedical Optics Express*, 3(10), 2611–2622.
- Skavenski, A. A., Hansen, R. M., Steinman, R. M., & Winterson, B. J. (1979). Quality of retinal image stabilization during small natural and artificial body rotations in man. *Vision Research*, 19(6), 675–683.
- Stevenson, S. B., Sheehy, C. K., & Roorda, A. (2016). Binocular eye tracking with the tracking scanning laser ophthalmoscope. *Vision Research*, 118, 98–104.
- Tulunay-Keesey, Ü. (1960). Effects of involuntary eye movements on visual acuity. *Journal of the Optical Society of America*, 50(8), 769–774.
- Tulunay-Keesey, Ü., & Bennis, B. J. (1979). Effects of stimulus onset and image motion on contrast sensitivity. *Vision Research*, 19(7), 767–774.
- Tulunay-Keesey, Ü., & Jones, R. M. (1976). The effect of micromovements of the eye and exposure duration on contrast sensitivity. *Vision Research*, 16(5), 481–488.
- Tulunay-Keesey, Ü., & Jones, R. M. (1980). Contrast sensitivity measures and accuracy of image stabilization systems. *Journal of the Optical Society of America*, 70(11), 1306–1310.
- Watanabe, A., Mori, T., Nagata, S., & Hiwatashi, K. (1968). Spatial sine-wave responses of the human visual system. *Vision Research*, 8(9), 1245–1263.
- Watson, A. B. (2014). A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision*, 14(7):15, 1–17, <https://doi.org/10.1167/14.7.15>. [PubMed] [Article]
- Watson, L. M., Strang, N. C., Scobie, F., Love, G. D., Seidel, D., & Manahilov, V. (2012). Image jitter enhances visual performance when spatial resolution is impaired. *Investigative Ophthalmology and Visual Science*, 53(10), 6004–6010.
- Wehrhahn, C. (2011). Psychophysical and physiological evidence contradicts a model of dynamic image stabilization. *Proceedings of the National Academy of Sciences, USA*, 108(10), E35–E35.
- Yang, Q., Arathorn, D. W., Tiruveedhula, P., Vogel, C. R., & Roorda, A. (2010). Design of an integrated hardware interface for AOSLO image capture and cone-targeted stimulus delivery. *Optics Express*, 18(17), 17841–17858.
- Yarbus, A. L. (1967). Perception of objects stationary relative to the retina. In *Eye Movements and Vision* (pp. 59–101). Boston, MA: Springer US.
- Zuber, B. L., Stark, L., & Cook, G. (1965, December 10). Microsaccades and the velocity-amplitude relationship for saccadic eye movements. *Science*, 150(3702), 1459–1460.